Measurement of JND Scales for Digital Video Sequences

Report prepared for the the IEEE G-2.1.6 Subcommittee on Video Compression Measurements meeting on Monday, August 7th, 2000 at Motorola at Schaumberg, IL.

Andrew B. Watson
NASA Ames Research Center
Moffett Field, CA, 94035-1000
abwatson@mail.arc.nasa.gov
http://vision.arc.nasa.gov/

Introduction

The purpose of this report is provide the IEEE G-2.1.6 Subcommittee on Video Compression Measurements with an update and some preliminary results from an ongoing study of video image quality. This report follows an earlier report to that committee which described general concepts, methods, and plans. Readers are encouraged to consult that report for additional details (http://grouper.ieee.org/groups/videocomp/ or http://vision.arc.nasa.gov/jnd/).

Basic Concepts

First we remind the reader of essential concepts and terminology. Each instance of a video sequence is assumed to induce in the observer a magnitude of perceived artifact. The perceived magnitudes are assumed to have unit variance. A scale of physical intensity, such as the amount of the added artifact, yields a scale of perceived artifact, in units of JND (one standard deviation). We call this the *scale function* or *JND function.* In experiments that vary the weight of the linear combination of an original and a compressed version of a video sequence, the physical measure of the artifact is the linear weight (also sometimes called *strength*). The top of this scale (weight=1) has a JND value that we call the total JND, and is the measure of JND for that compressed condition (HRC & SRC). The goal of this project is to estimate the JND scales and total JNDs for a number of the conditions used in the VQEG experiments.

The EASE Method

In the previous report, we proposed using the Method of Concatenated Thresholds (MCT) to piece together the JND scale from a set of successive measurements of discrimination thresholds. We also suggested as a further project the development of an efficient method for estimating the scale. Since that report, we determined that the MCT was problematic, and went forward with the development of an efficient method. We have given this new method the name EASE (Efficient Adaptive Scale Estimation). Here we give a brief description of this method.

EASE proceeds as pictured in Table 2. A mathematical form is assumed for the scale function. Initial parameters are guessed. Based on the function, a set of pairs of weights are selected that are approximately a specified number of JNDs apart. One trial is then conducted at each of those pairs. All of the data are then used to estimate the function parameters. If sufficient data have been collected, the procedure stops, and the data may the be used to estimate the JNDs for the condition. If more trials are needed, a new set of pairs are derived from the newly estimated function, and the procedure continues.

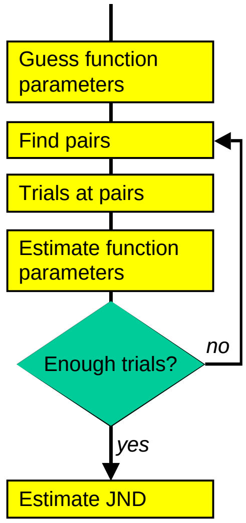

Figure 1. Flow diagram for the EASE method.

Various mathematical functions could be assumed for the scale function within the EASE method. Typically, the function will have one parameter that is the value of the scale when the weight is 1 - in other words, the total JND . We will call this generic parameter *max*.

The termination rule we have used is to test whether trials/max >= trials_per_jnd. We have used a value of trials_per_jnd = 32. The rationale is that we believe that about 32 trials per JND are required to obtain sufficient accuracy.

To examine this issue, we have conducted some simulations of the EASE procedure. Figure 2 shows the standard deviation in the estimate of max, derived from 100 replications, of an EASE session with the value of trials/JND shown on the abscissa. At 32 trials/JND, a reasonable compromise between accuracy and time is obtained.

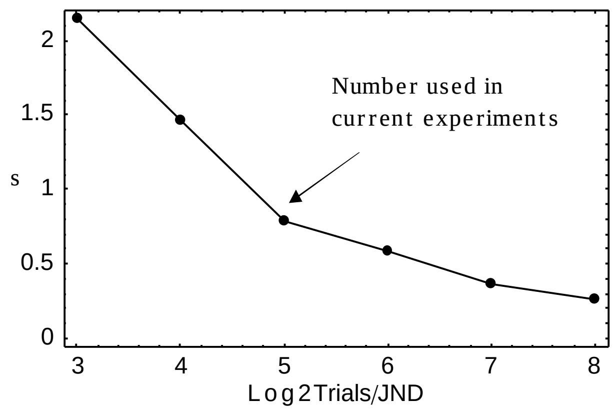

Figure 2. Effect of trials/JND on accuracy of the EASE procedure.

The number of trials also has an effect on the bias of the estimate. As shown in , the estimates are always biased slightly upwards, and this bias is reduced as the number of trials increases. For the trials/JND criterion used in these experiments, a bias of as much as 0.5 JND may be present.

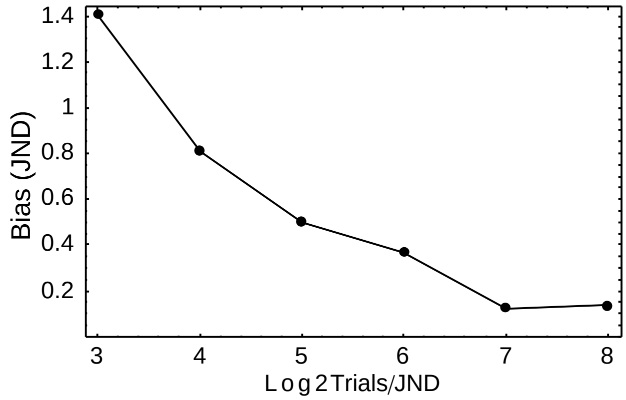

Figure 3. Effect of trials/JND on bias in the EASE procedure.

We should note that these simulations consider only a small set of possible conditions, and that a more complete examination of the statistical properties of the procedure would be valuable.

Estimating JND from EASE Data

Once an EASE session is complete, the data can be analyzed to derive a scale function, and an estimate of the total JND for the condition. We have used two methods of analysis.

Piecewise estimation of the scale function

In this method we fit a function that is defined piecewise by its values at each of the weight values used in the session. The parameters are simply the values of the function at those points (more precisely, their differences). The parameters are estimated by minimizing the negative log of the likelihood function. The total JND is simply the last point on the function, at a weight of 1. We call this estimate *tot*. An example of the results of this procedure for one session is shown in Figure 4. Each point in the figure corresponds to a weight used in the session. The vertical coordinate is the estimated number of JNDs (from weight = 0) at that point.

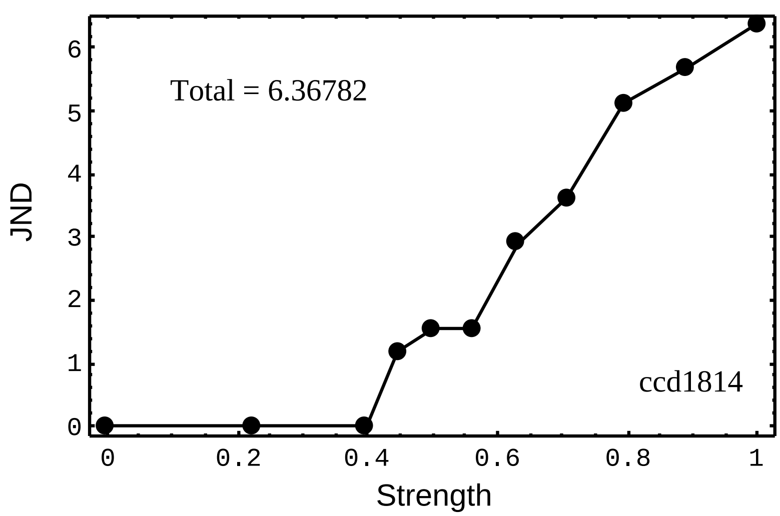

Figure 4. Piecewise estimate of the scale function.

Fitting a scale function

In this method we assume a particular mathematical parameterized form for the scale function, and fit it to the data, again using a maximum likelihood method. The function we have used is a truncated power function:

$$\mathtt{max}\,(1. - \mathtt{threshold})^{-\mathtt{power}}\,\mathtt{Max}[0, -\mathtt{threshold} + \mathtt{w}]^{\mathtt{power}} \qquad (1$$

where w is the weight, max is the total JNDs, threshold is the weight below which the scale remains at zero, and power is the exponent of the power function. This function was in part inspired by results such as those in , and also because this function is widely used in psychophysical scaling theory(Engen, 1972; Falmagne, 1986). The first term is present to ensure that the parameter max is the value of the function at 1 (total JND).

The results of applying this method to one session are shown in Figure 5, along with the previous piecewise estimates. In this case, we see very good agreement between the two methods, both in the overall course of the function, and in its asymptote (the total JND). The estimated parameters are also shown in the figure.

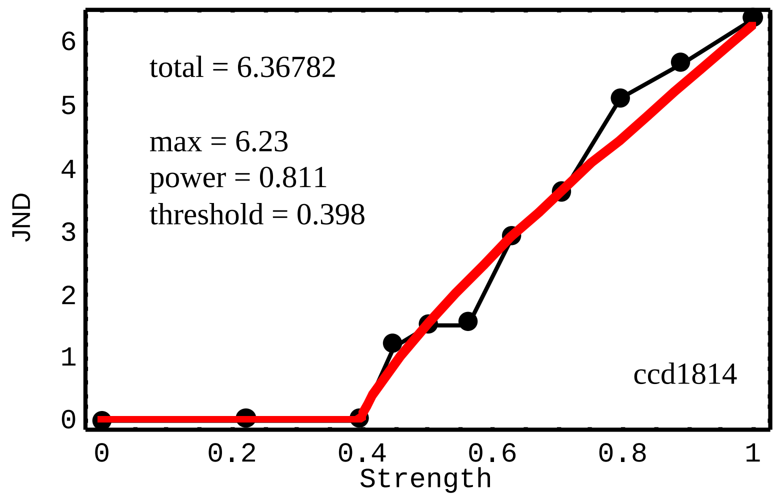

Figure 5. Functional estimate of the scale function.

Agreement between the two methods, both in total JND and in the shape of the JND function, is not always as good as that in . but on average the function adopted (Eq. 1) appears to be reasonable.

Failures of Piecewise Method

It is important to note a potential problem with the piecewise method. Consider the results in Figure 6. Here we see a gross disagreement between the two estimation methods (32.3 vs 9.35). We also see that several of the increments in the piecewise function are as large as 5 JND. The problem is that a step that large would yield essentially perfect performance, but perfect performance is consistent with any large number of JNDs. That is why the ease procedure tries to arrange pairs that are about 1 JND apart. This situation typically arises when the observer has guessed correctly on the few trials (perhaps only one) that have been presented at that separation. In short, the piecewise estimates are under-constrained and unreliable when there are too few trials. In this example, there were 99 trials, or trials/max ≈ 10, well below our recommended value of 32.

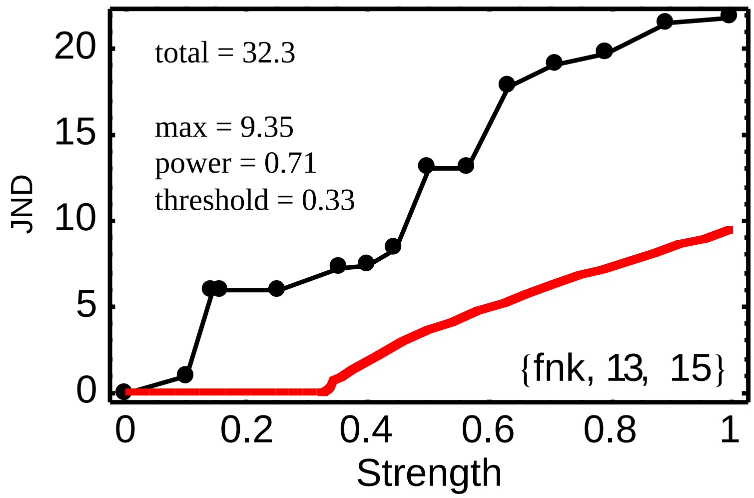

Figure 6. Example of under-constrained piecewise estimate.

Experiment 1

Experiment 1 follows the general outlines of our earlier proposal, except that the EASE method is used and the number of conditions is altered.

Conditions

We used a total of 25 conditions. These consisted of the complete matrix obtained by combining 5 SRCs with 5 HRCs. All conditions employed 60 Hz 525 line video.

The selected SRCs are shown in Figure 7, along with the identifying number used in the VQEG study(VideoQualityExpertsGroup, 2000).

The HRCs are indicated in Table 1. This table also shows the mean DMOS score obtained by each HRC in the VQEG experiment. These values were used to select a set of HRCs that spanned the quality range employed in the VQEG experiments.

Figure 7. SRCs used in Experiment 1.

HRC	Mean DMOS	Mbps	CODEC	Details
7	5.82	6	mp@ml	
5	13.83	8 & 4.5	mp@ml	Two codecs concatenated
9	21.18	3	mp@ml	
14	33.35	2	mp@ml	Horizontal resolution reduction
15	45.75	0.768	H.263	CIF, Full Screen

Table 1. HRCs used in Experiment 1.

Methods

Video sequences were presented under computer control and displayed on a studio quality television monitor capable of displaying ITU-601 digital video streams. The display apparatus consists of an SGI Octane computer with SDI serial digital video input/output board, a Ciprico FibreChannel Disk Array, and a SONY BVM 20E1U monitor.

Trials were conducted using the EASE procedure in blocks of 32 trials. Ease parameters were trials_per_jnd = 32, and jnd_step = 1.273. In some sessions, the criterion number of trials was never reached. We expect to factor in the number of trials into the final data analysis.

Each presentation was seven seconds in duration, comprising the first seven seconds of each vqeq sequence. Viewing conditions followed Rec. 500. Viewing distance was 3H, 5H or 7H. Observers were checked for normal color vision, and normal or corrected-to-normal acuity. We also recorded the age, gender, and handedness of the observers.

Results

We have completed at least one session for each of the 25 conditions at the 5H viewing distance. A small subset of data have been collected for the 3H and 7H viewing distances. To date we have conducted 5983 trials, representing about 35 hours of observation time.from a total of 34 observers.

Data for SRC 13 are shown in Figure 8. Where error bars are shown, they indicate ±1sd. Where not shown, they indicate only one estimate has thus far been obtained.

The principal message to be drawn from this figure at this time is that we have shown that it is possible to measure JNDs with some reliability for various HRCs. The JND values are also of interest. The results show JND values ranging between around 2 and 10, depending upon HRC.

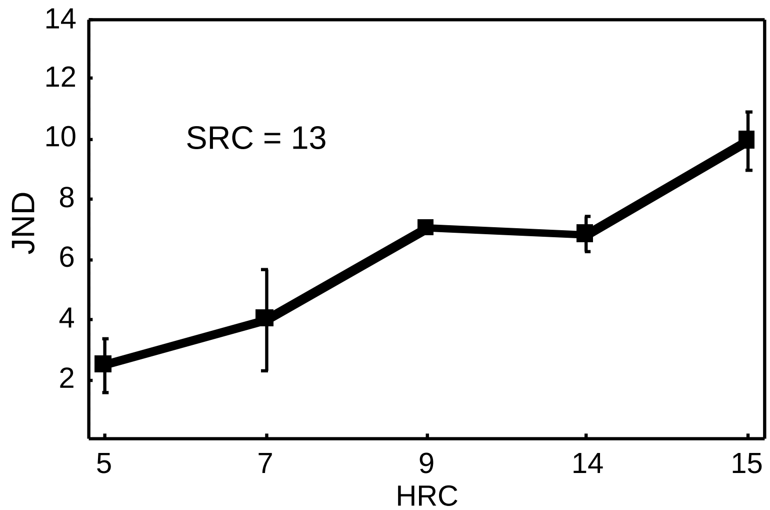

Figure 8. Total JNDs for SRC 13 at five HRCs.

Comparing "tot" and "max"

As noted above, we have considered two methods of estimating the total JND measure for a particular condition. The first measure "tot" is obtained by estimation of the values of the JND function at each weight used, while the second measure "max" is a parameter of a fitted function. Although "tot" is a more "model'free" measure, we have noted earlier a potential problem. When too few trials/JND are collected, the value of tot is biased upwards.

To compare estimated values of max and tot they are plotted against one another in Figure 9. Despite at least two instances of this upward bias, in most cases the two measures give nearly identical results. This suggests that we should use max as our definitive measure.

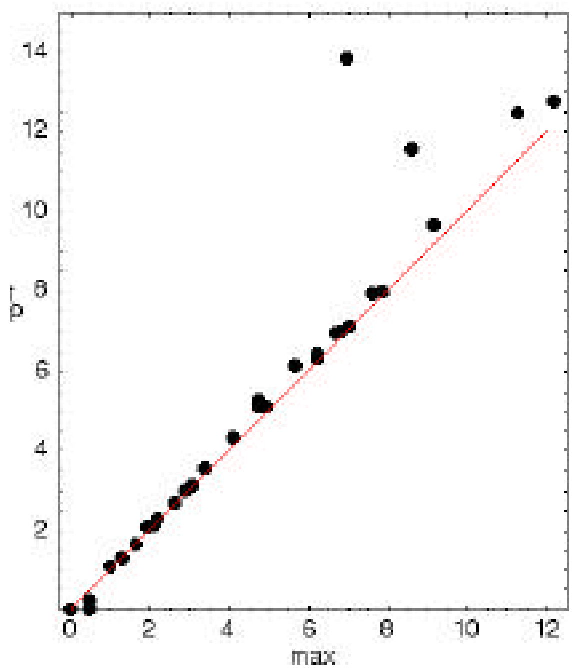

Figure 9. Comparison of max and tot, two measures of the total JND for a given condition.

Variability

The repeatability of our JND measurements is an obvious concern. A preliminary assessment of this can be obtained by looking at the standard deviation of our measurements, for those conditions where we have collected at least three replications. These are shown in Table 2. The average standard deviation of 0.82 JND is acceptably small. It should be noted that this figure includes any between-observer variability, since the three replications are from three different observers.

SRC	HRC	sd
13	5	0.872646
13	14	0.579791
13	15	0.959998
18	14	0.890953
	mean	0.825847

Table 2. Standard deviation (MLE) for conditions with three replications.

JND vs MDOS

Although our data are preliminary and incomplete, it is of interest to compare our JND measurements with the MDOS (mean differential opinion scores) measurements obtained by VQEG. Again, error bars indicate ±1sd where more than one estimate is available. The Pearson correlation between the two measures is 0.86. This agreement between the two varieties of measurement is reasuring, and suggests that JND measurements are at least as valid as DSCQS measurements of video quality. The line shows the best fitting linear relation, which has a slope of about 0.215. This means that the DSCQS measurements are about five times the corresponding number of JNDs. This proportionality would not be expected to persist, of course, if DSCQS measurements are subject to context effects.

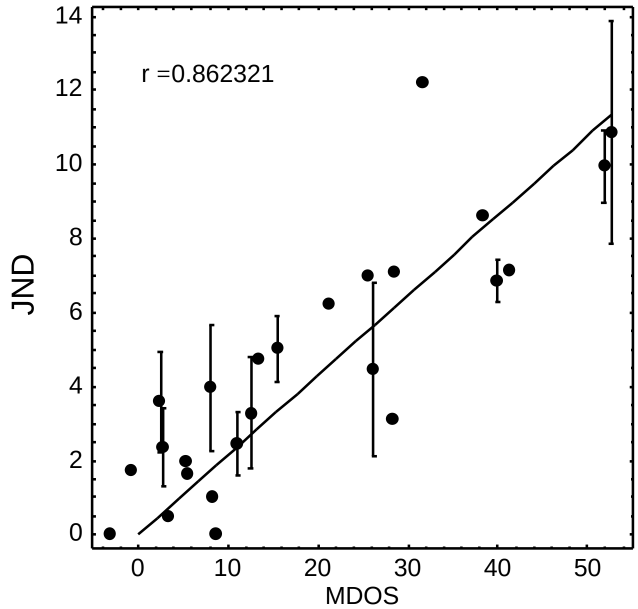

Figure 10. JND versus MDOS.

Effect of Viewing Distance

Viewing distance is one of the key determinants of visual quality. We selected a range of distances that would span the typical range of viewing conditions: 3H, 5H, and 7H (H =picture height). At present, we have only a small amount of data at the 3H and 7H distances. In Figure 11, we see an expected decline in JND as viewing distance increases. Additional data will be required to determine the quantitative nature of this decline, and how much it varies with SRC and HRC.

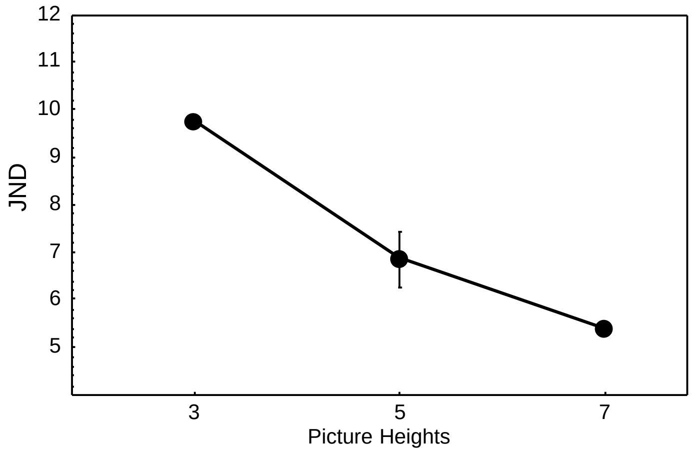

Figure 11. Effect of viewing distance on JND.

Applications of JND Measurements

The JND measurements collected in this project could serve three purposes.

1. The first is in the creation of standard samples of video with specified JND levels. These could serve as a basis for visual comparison in a variety of video production and distribution applications.
2. The second purpose is as a dataset for testing of video quality metrics. An advantage of this dataset is that it specifies subjective artifacts in absolute measures (JNDs) and that it contains multiple viewing distances.
3. The third purpose is as a basis for calibration of existing and future video quality metrics. At present, these metrics yield arbitrary units which have only relative meaning. Through correlation with JND measures, they can be converted to absolute measures.

Summary

1. The EASE method appears to be a reliable method for measuring JNDs for processed video sequences.
2. Substantial data have been collected at the 5H viewing distance on the selected 25 conditions.
3. A small amount of data have been collected at 3H and 7H distances.
4. Preliminary analyses have been made of variability, correlation with VQEG DMOS scores, and effects of SRC, HRC, and viewing distance.
5. JND measurements offer a means creating calibrated artifact samples, and of testing and calibrating video quality models.
6. Future work will include completing data collection and comparison of JNDs and metric results.

References

Engen, T. (1972). Psychophysics II. Scaling Methods. In J. W. Kling, & L. A. Riggs (Ed.), Woodworth and Schlosberg's Experimental Psychology (pp. 47-86). New York: Holt, Rinehart and Winston.

Falmagne, J. C. (1986). Psychophysical measurement and theory. In K. Boff, L. Kaufman, & J. Thomas (Ed.), Handbook of Perception and Human Performance (pp. 1.3-1.66). New York: Wiley.

VideoQualityExpertsGroup. (2000). Final report from the video quality experts group on the validation of objective models of video quality assessment http://www.crc.ca/vqeg/.